Sonic Interaction Design to enhance presence and motion in virtual environments

Nordahl, Rolf

Published in:
Proceedings of CHI 2008 Workshop on Sonic Interaction Design

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Sonic Interaction Design to enhance presence and motion in virtual environments

Rolf Nordahl
Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750
rn@media.aau.dk

ABSTRACT
An occurring problem of the image-based-rendering technology for Virtual Environments has been that subjects in general showed very little movement of head and body since only visual stimulus was provided. By transferring information from film studies and current practice, practitioners emphasize that auditory feedback such as sound of footsteps signifies the character giving them weight and thereby subjecting the audience to interpretation of embodiment.

We hypothesize that the movement rate can be significantly enhanced by introducing auditory feedback. In the described study, 126 subjects participated in a between-subjects experiment involving six different experimental conditions, including both uni and bi-modal stimuli (auditory and visual). The aim of the study was to investigate the influence of auditory rendering in stimulating and enhancing subjects motion in virtual reality. The auditory stimuli consisted of several combinations of auditory feedback, including static sound sources as well as self-induced sounds. Results show that subjects' motion in virtual reality is significantly enhanced when dynamic sound sources and sound of egomotion are rendered in the environment.

Author Keywords
Interactive auditory feedback, physical models, presence

ACM Classification Keywords
H 5.5 Sound and Music Computing ; H 5.2 User Interfaces

INTRODUCTION
In the realm of Virtual Reality (VR) and Virtual Environments (VE) sound has not until very recently been considered of value when one wishes to mediate experiences to the participant. Although sound is one of the fundamental modalities in the human perceptual system, it still contains a large area for exploration for researchers and practitioners of VR [15]. While research has provided different results concerning multimodal interaction among the senses, several questions remain in how one can utilize e.g., audiovisual phenomena when building interactive VR experiences.

Following the computational capabilities of evolving technology, VR-research has moved from being focused on unimodality (e.g. the visual modality) to new ways to elevate the perceived feeling of being virtually present and to engineer new technologies that may offer a higher degree of immersion, here understood as presence as immersion [9].

Engineers have been interested in the audio-visual interaction from the perspective of optimizing the perception of quality offered by technologies [6, 14]. Furthermore, studies have shown that by utilizing audio, the perceived quality of lower quality visual displays can increase [16].

Likewise researchers from neuroscience and psychology have been interested in the multimodal perception of the auditory and visual senses [8]. Studies have been addressing issues such as how the senses interact, which influences they have on each other (predominance), and audio-visual phenomena such as the cocktail party effect [2] and the ventriloquism effect [7].

Among the initiatives to investigate how technology can enhance sense of immersion in virtual environments, the recently completed BENOGO project[1] had as its main focus the development of new synthetic image rendering technologies (commonly referred to as Image Base Rendering (IBR)) that allowed photo-realistic 3D real-time simulations of real environments. The project aimed at providing a high degree of immersion to subjects for perceptual inspection through artificial created scenarios based on real images. Throughout the project the researchers wished to contribute to a multilevel theory of presence and embodied interaction, defined by three major concepts: immersion, involvement and fidelity.

One of the drawbacks of reconstructing images using the IBR technique is the fact that, when the pictures are captured, no motion information can be present in the environment. This implies that the reconstructed scenarios as static over time. Depth perception and direction are varied according to the motion of the user, which is able to investigate the environment at 360^0 inside the so-called region of exploration (REX). However, no events happen in the environment, which make it rather uninteresting to explore [11].

In this paper we advocate the use of interactive auditory feedback as a mean to enhance immersion in a photorealistic virtual environment. We focus both on ambient sounds, defined as sounds characteristic of a specific environment which the user cannot modify, as well as interactive sounds, which were synthesized in real-time and controlled by actions of users in the environment. Such sounds were driven by using a footsteps controller able to capture the motion of

[1] www.benogo.dk

users in the environment, and subsequently produce in real-time sounds produced by walking on different surfaces.

Furthermore, sounds were spatialized in a 8-channels surround sound system utilizing the Vector Based Amplitude Panning (VBAP) algorithm [12]. Our hypothesis is that augmenting the environment with interactive sounds will enhance motion of the subjects. Measuring the quantity of motion is important since we hypothesize that a higher level of motion implies that subjects explore the environment more actively and therefore with an increased interest.

A MULTIMODAL ARCHITECTURE

Figure 1 shows a schematic representation of the multimodal architecture used for the experiments. The visual stimulus was provided by a standard PC running Suse Linux 10. This computer was running the BENOGO software using the REX disc called *Prague Botanical Garden*.

The Head-Mounted-Display (HMD) used was a VRLogic V82. It features Dual 1.3 diagonal Active Matrix Liquid Crystal Displays with resolution per eye: ((640x3)x480), (921,600 color elements) equivalent to 307,200 triads. Furthermore the HMD provides a field of view of 60^0 diagonal. The tracker used was a Polhemus IsoTrak II3. It provides a latency of 20 milliseconds with a refresh rate of 60 Hz.

The audio system was created using a standard PC running MS Windows XP SP 2. All sound was run through Max/MSP[2] and as output module a Fireface 800 from RME5[3] was used. Sound was delivered by eight Dynaudio BM5A speakers[4]. Figure 2 shows a view of the surround sound lab where the experiments were run. In the center of the picture, the tracker's receiver is shown.

AUDITORY RENDERING

In the laboratory eight speakers were positioned in a parallelepipedal configuration. Current commercially available sound delivery methods are based on sound reproduction in the horizontal plane. However, we decided to deliver sounds in eight speakers and thereby implementing full 3D capabilities. By using this method, we were allowed to position both static sound elements as well as dynamic sound sources linked to the position of the subject. Moreover, we were able to maintain a similar configuration to other virtual reality facilities such as CAVEs[5], where eight channels surround is presently implemented. This is the reason why 8-channels sound rendering was chosen compared to e.g., binaural rendering [3].

As described, two computers were installed in the laboratory, one running the visual feedback described in the following section, and one running the auditory feedback. A Polhemus tracker, attached to the head mounted display, was connected to the computer running the visual display, and allowed to track the position and orientation of the user in 3D. The computer running the visual display was connected to

[2]www.cycling74.com

[3]http://www.rmeaudio.com/english/firewire/

[4]http://www.dynaudioacoustics.com

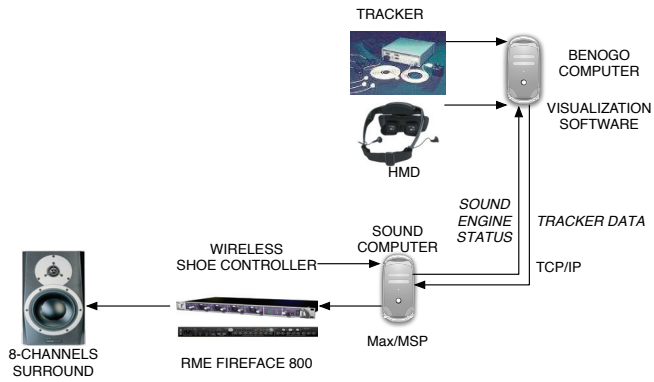

Figure 1. Connection of the different hardware components in the experimental setup.

the computer running the auditory display by TCP/IP. Connected to the sound computer there was the interface RME Fireface 800 which allowed delivering sound to the eight channels, and the wireless shoe controller. The mentioned controller, developed specifically for these experiments [10], allowed detecting the footsteps of the subjects and mapping these to the real-time sound synthesis engine. The different hardware components are connected together as shown in Figure 1.

Four kinds of auditory feedback were provided to the subjects:

1. "Static" soundscape, reproduced at max. peak of 58dB, measured c-weighted with slow response. This soundscape was delivered through the 8-channels system.
2. Dynamic soundscape with moving sound sources, developed using the VBAP algorithm, reproduced at max. peak of 58dB, measured c-weighted with slow response.
3. Auditory simulation of ego-motion, reproduced at 54 dB. (This has been recognised as the proper output level as described in [11])
4. A piece of classic music as described before, reproduced at max. peak of 58dB, measured c-weighted with slow response.

However, six testing conditions were implemented, as described later.

Interactive auditory feedback

A real-time footstep synthesizer, controlled by the subjects using a set of sandals embedded with pressure sensitive sensors was designed. By navigating in the environment, the user controlled synthetic sounds. Footsteps recorded on seven different surfaces were obtained from the Hollywood Edge Sound Effects library.[5] The surfaces used were metal, wood, grass, bricks, tiles, gravel and snow. The sounds were analyzed and the analysis results used to build a footsteps synthesizer using a combination of modal synthesis [1] and phys-

[5]www.hollywoodedge.com

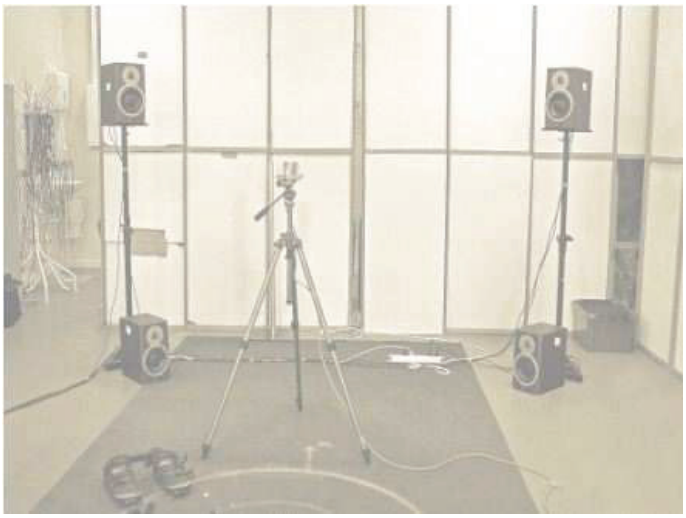

Figure 2. A view of the 8-channels surround sound lab where the experiments were run.

ically informed stochastic models (PHYSM) [4]. More specifically, regular surfaces such as bricks, metal, wood and tiles were synthesized using a modal synthesizer with few (two or three) resonances. Grass, gravel and snow were synthesized using the PHYSM algorithm. In order to control the synthetic footsteps in the virtual reality environment, subjects were asked to wear a pair of sandals embedded with pressure sensitive sensors placed one in each heel. Such sandals are shown in Figure 3.

Despite its simplicity, the shoe controller was effective in enhancing the user's experience, as it will be described later. While subjects were navigating around the environment, the sandals were coming in contact with the floor, thereby activating the pressure sensors. Through the use of a microprocessor, the corresponding pressure value was converted into an input parameter which was read by the real-time sound synthesizer Max/MSP.[6] The sensors were wirelessly connected to a microprocessor, as shown in Figure 3, and the microprocessor was connected to a laptop PC.

The continuos pressure value was used to control the force of the impact of each foot on the floor, to vary the quality of the synthetic generated sounds. The use of physically based synthesized sounds allowed to enhance the level of realism and variety compared to sampled sounds, since the produced sounds of the footsteps depended on the impact force of subjects in the environment, and therefore varied dynamically. The different simulated surfaces were activated according to the virtual place which the users were visiting, and rendered through and 8-channel surround sound system.

VISUAL FEEDBACK

The visual feedback used in these experiments was created under the BENOGO project. The idea behind this project is the creation of photorealistic visual environments obtained by taking pictures of a specific location at different angles,

[6] www.cycling74.com

Figure 3. The sandals enhanced with pressure sensitive sensors wirelessly connected to a microprocessor.

and building a reconstruction of the same place at the computer using image based rendering techniques. In this specific experiment, subjects were looking at pictures from the Prague botanical garden, which is shown in Figure 4.

One of the peculiarities of this approach is the fact that no moving objects have to be present in the environment when the pictures are taken, since this would affect the visual reconstruction. This also implies that the reconstructed scenarios do not vary over time, which means that one could be concerned with that the exposure to the environment becomes tedious and uninteresting for the users to explore. As such, we regard the exploration of auditory feedback as a good way to cope with these limitations, as explained in the following section.

Figure 4. An image of the Prague botanical garden used as visual feedback in the experiments.

TEST DESCRIPTION

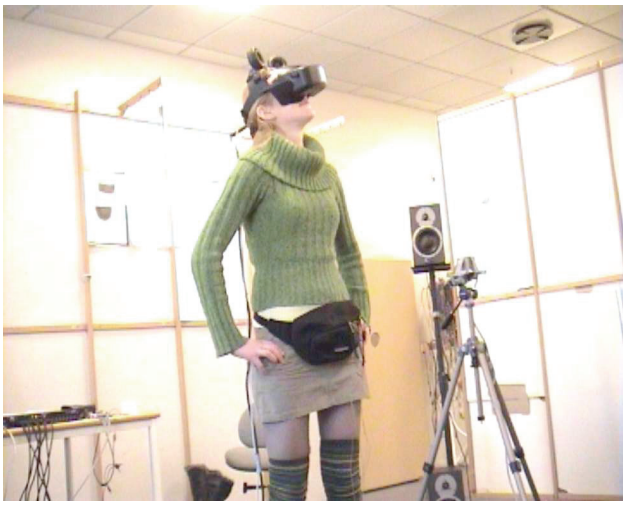

Figure 5. A subject navigating in the virtual environment wearing an head mounted display HMD)

126 subjects took part to the experiment. All subjects reported normal hearing and visual conditions. Figure 5 shows one of the subjects participating to the experiment. Before entering the room, subjects were asked to wear a head mounted display and the pair of sandals enhanced with pressure sensitive sensors. Subjects were not informed about the purpose of the sensors-equipped footwear. Before starting the experimental session the subjects were told that they would enter a photo-realistic environment, where they could move around if they so wished. Furthermore, they were told that afterwards they would have to fill out a questionnaire, where several questions would be focused on what they remember having experienced. No further guidance was given.

The experiment was performed as a between subjects study including the following six conditions:

1. Visual only. This condition had only uni-modal (visual) input.

2. Visual with footstep sounds. In this condition, the subjects had bi-modal perceptual input (audio-visual) comparable to our earlier research [11].

3. Visual with full sound. This condition implies that subjects were treated with full perceptual visual and audio input. This condition included static sound design, 3D sound (using the VBAP algorithm) as well as rendering sounds from ego-motion (the subjects triggered sounds via their footsteps).

4. Visual with full sequenced sound. This condition was strongly related to condition 3. However, it was run in three stages: the condition started with bi-modal perceptual input (audio-visual) with static sound design. After 20 seconds, the rendering of the sounds from egomotion was introduced. After 40 seconds the 3D sound started (in this case the sound of a mosquito, followed by other environmental sounds).

5. Visual with sound + 3D sound. This condition introduced bi-modal (audio-visual) stimuli to the subjects in the form of static sound design and the inclusion of 3D sound (the VBAP algorithm using the sound of a mosquito as sound source). In this condition no rendering of ego-motion was conducted.

6. Visual with music. In this condition the subjects were introduced to bi-modal stimuli (audio and visual) with the sound being a piece of music[7] described before. This condition was used as a control condition, to ascertain that it was not sound in general that may influence the in- or decreases in motion. Furthermore it enabled us to deduce if the results recorded from other conditions were valid. From this it should be possible to deduct how the specific variable sound design from the other experimental conditions affects the subjects.

CONDITION	AUDITORY STIMULI	NUM SUBJ	MEAN (AGE)	ST.D. (AGE)
Visual	None	21	25.6	4.13
Visuals w. foot	3	21	25.7	3.75
Full	1 + 2 +3	21	25	4.34
Full seq	1 + 2 + 3	21	22.8	2.58
Sound + 3D	1+2	21	22.9	2.5
Music	4	21	28	8.1

Table 1: Description of the six different conditions to which subjects were exposed during the experiments. The number in the second column refers to the auditory feedback previously described.

RESULTS

Tracked movement	Mean	Median	St.d.
Visual only	21.41	21.61	6.39
Visual w. foot	22.82	25.66	6.89
Full	26.47	26.54	5.6
Full Seq	25.19	24.31	5.91
Sound + 3D	21.77	21.87	6.74
Music	20.95	20.79	6.39

Table 2: Motion analysis for the different conditions.

	Visual only	Visual w. foot	Full	Full seq.	Music
Visual only			0.006	0.03	0.41
Visual w. foot	0.26		0.04	0.132	0.197
Full					
Full seq.			0.243		0.018
Sound + 3D	0.431	0.32	0.022	0.048	0.347
Music			0.003		

Table 3: Comparison of the motion analysis for the different conditions (p-value).

[7]Mozart, Wolfgang Amadeus, Piano Quintet in E flat, K. 452, 1. Largo Allegro Moderato, Philips Digitals Classics, 446 236-2, 1987

Table 2 shows the results obtained by analysing the quantity of motion over time for all subjects for the different conditions. Such analysis was performed by calculating motion over time using the tracker data, where motion was defined as Euclidian distance over time for the motion in 3D. Since motion was derived from the tracker's data placed on top of the head mounted display, only the motion of the head of the subjects was tracked, and not additional body motion.

The significance of the results is outlined in Table 3, where the corrected p-value was calculated for the different conditions, using a t-test. As can be seen from Table 3, there exists a clear connection between the stimuli. First of all it is interesting to notice that the condition *Music* elicits the lowest amount of movement, even less than the condition *Visual Only*. However, the difference between the condition *Visual Only* and *Music* is not significant (p=0.410), which translates into that we cannot state that using sounds not corresponding to the environment (such as music), should diminish the amount of movement. The fact that music shows less movement indicates that it is important which sound is used. The condition *Music* was in fact used as control condition for this very purpose. Results also show that footsteps sounds alone do not appear to cause a significant enhancement in the motion of the subjects. When comparing the results of the conditions *Visual only* versus *Visuals w. footsteps* (no significant difference) and the conditions *Full* versus *Sound+3D* (significant difference) there is an indication that the sound of footsteps benefits from the addition of environmental sounds. This result shows that environmental sounds are implicitly necessary in a virtual reality environment and we assume that their inclusion is important to facilitate motion.

Figure 6 shows the visualization of the Polhemus tracker data for one subject with visual only stimuli (top) and with full condition (bottom). The increase of movement in the full condition is clearly noticeable.

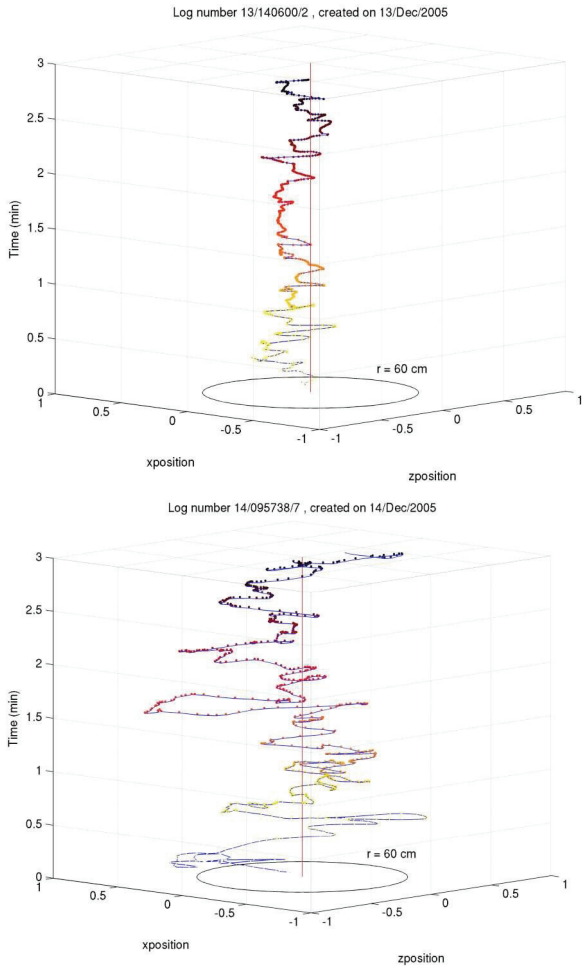

Figure 6. Top: visualization of the motion of one subject over time with visual only condition (top) and full condition (bottom).

MEASURING PRESENCE

As a final analysis of the six experimental conditions, we investigated the qualitative measurements of the feeling of Presence. Through the tests for all conditions we implemented all questions from the SVUP-questionnaire [17]. The SVUP is concerned with examining 4 items, where the most important item in relation to our thesis is the feeling of Presence. The SVUP-questionnaire does so by asking the subjects to answer 4 questions which all relates to the feeling of presence. The results of these answers are then averaged for each subject, resulting in what is referred to as the presence index. The questions related to the naturaleness of interaction with the environment, and sense of presence and involvement in the experience. All answers were given on a Likert-scale [10], from 1-7, (from 1 represents not at all, and 7 represents very much).

Presence index	Mean	Median	St.d.
Full	4.77	4.75	1.08
Music	4.82	5	1.13
Full Seq	4.79	4.75	0.69
Visual only	4.58	4.5	0.92
Visual w. foot	4.82	5	1.06
Sound + 3D	4.81	5	0.79

Table 4: Average presence index for the 6 experimental conditions.

Table 4 shows the results of the presence questionnaire for the different conditions. As can be seen from the Table, no significant differences were noticeable among the different conditions.

One reason that may affect the overall results derived from the self-report of the subjects is that the experiments of this thesis were done as a between-subjects exploratory study. The fact that the individual subject only experienced one condition was optimal in the sense that issues concerning

subjects becoming accustomed to the VE or finding it increasingly boring was minimized.

However, since the subjects have no other conditions as a frame of reference, this may be a plausible cause of what we have experienced through these results of the SVUP presence index, i.e., that between-subjects as a method for this particular presence index is not adequate since the subjects give their initial feeling of how they felt without having anything to measure this feeling against. However, the quantitative data from the motion tracking shows clear results with significance and the between-subjects strategy is well suited towards that such experiments.

CONCLUSION

In this paper we investigated the role of dynamic sounds in enhancing motion and presence in virtual reality. Results show that 3D sound with moving sound sources and auditory rendering of ego-motion significantly enhance the quantity of motion of subjects visiting the VR environment.

It is very interesting to notice that it is not the individual auditory stimulus that affects the increase of motion of the subjects, but rather that it is the combination of soundscapes, 3-dimensional sound and auditory rendering of ones own motion that induces a higher degree of motion.

We also investigated if the sense of presence was increased when interactive sonic feedback was provided to the users. Results from the SVUP presence questionnaire do not show any statistical significance in the increase of presence.

We are currently extending these results to environments were the visual feedback is more dynamic and interactive, such as computer games and virtual environments reproduced using 3D graphics.

REFERENCES

1. J.M. Adrien. The missing link: modal synthesis. *Representations of musical signals table of contents*, pages 269–298, 1991.
2. B. Arons. A review of the cocktail party effect. *Journal of the American Voice*, 12, 1992.
3. D.R. Begault. *3-D sound for virtual reality and multimedia*. AP Professional Boston, 1994.
4. P.R. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.
5. C. Cruz-Neira, D.J. Sandin, T.A. DeFanti, R.V. Kenyon, and J.C. Hart. The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.
6. N.F. Dixon and L. Spitz. The detection of audiovisual desynchrony. *Perception*, 9, 1980.
7. S. Handel and I. NetLibrary. *Perceptual Coherence Hearing and Seeing*. Oxford University Press, 2006.
8. A Kohlrausch and S. Vand de. Par. Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications. In *Proceeding of IST/SPIE Conference on Human Vision and Electronic Imaging IV*, 1999.
9. M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):20, 1997.
10. TJ MAURER and HR PIERCE. A comparison of Likert scale and traditional measures of self-efficacy. *Journal of applied psychology*, 83(2):324–329, 1998.
11. R. Nordahl. Auditory rendering of self-induced motion in virtual reality. *M. Sc. project report, Dept. of Medialogy, Aalborg University Copenhagen*, 2005.
12. V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.
13. V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Helsinki University of Technology, 2001.
14. S. Rihs. The influence of audio on perceive picture quality and subjective audio-video delay tolerance. In *Proceeding of the MOSAIC workshop Advanced Methods for the Evaluation of Television Picture Quality*, 1995.
15. K.M. Stanney et al. *Handbook of virtual environments: design, implementation, and applications*. Lawrence Erlbaum Associates, 2002.
16. R.L. Storms and M.J. Zyda. Interactions in perceived quality of auditory visual displays. *Presence*, 9, 2000.
17. D. Västfjäll, P. Larsson, and M. Kleiner. Development and validation of the Swedish viewer-user presence questionnaire (SVUP), 2000.